

ALGORITMI GENETICI E MONTE CARLO: UN CONNUBIO PROMETTENTE PER L'OTTIMIZZAZIONE DI PROGETTO E GESTIONE LOGISTICA D'IMPIANTO

Marzio Marseguerra and Enrico Zio
Dipartimento di Ingegneria Nucleare, Politecnico di Milano
Via Ponzio 34/3, 20133 Milano, Italy
marzio.marseguerra@polimi.it

SOMMARIO

In questo lavoro viene presentato un approccio innovativo alla soluzione di problemi di ottimizzazione del progetto e della gestione di impianto, basato su una opportuna combinazione delle procedure di ricerca degli algoritmi genetici con tecniche di simulazione Monte Carlo.

1. INTRODUZIONE

Nella progettazione e gestione di un impianto a rischio occorre considerare le esigenze contrastanti di economicità e sicurezza; pertanto alle analisi quantitative di affidabilità devono affiancarsi le analisi dei costi e dei profitti. Ai guadagni di produzione si contrappongono i costi di acquisto dei macchinari, i costi di manutenzione e riparazione, i costi di indisponibilità del sistema, i costi di incidente. La progettazione diviene così un'attività tendente ad individuare i componenti e i parametri di gestione ottimali per quanto riguarda sia gli aspetti di sicurezza che la bilancia dei costi di operazione e benefici di produzione. Nella pratica l'elevato numero di parametri in gioco e le complesse non linearità presenti rendono la ricerca della configurazione di sistema ottimale difficilmente affrontabile con metodi analitici [1-4].

Il problema di ottimizzazione implica la valutazione delle prestazioni del sistema in diverse condizioni di progetto e di gestione logistica. A tal fine, occorre precisare il modello stocastico che descrive il comportamento affidabilistico del sistema, cioè i processi di guasto, riparazione e manutenzione dei suoi componenti. L'approccio proposto per la valutazione di tale modello prevede l'impiego di tecniche di simulazione Monte Carlo [5]. La flessibilità offerta dal metodo Monte Carlo consente di considerare molti aspetti pratici quali l'invecchiamento dei componenti, le corrispondenti manutenzioni periodiche atte a contrastarne gli effetti, le degradazioni che i componenti subiscono a seguito di riparazioni imperfette, la presenza di un limitato numero di squadre di riparazione e la loro gestione nel rispetto di opportune priorità di intervento assegnate ai componenti a seconda dell'importanza che essi rivestono nei confronti del funzionamento sicuro del sistema. Tuttavia, esiste un rovescio della medaglia nell'utilizzo del Monte Carlo in quanto i sistemi in studio sono tipicamente molto affidabili e quindi comportano la simulazione di numerose storie Monte Carlo al fine di ottenere risultati statisticamente significativi. Dunque, per ogni possibile soluzione del problema di ottimizzazione, cioè per ogni set di parametri di progetto e di logistica, occorre simulare migliaia di storie Monte Carlo. Questo tipo di procedura diventa impraticabile per sistemi reali.

Per ovviare a questo problema nel presente lavoro abbiamo esaminato la possibilità di utilizzare le procedure di ricerca degli algoritmi genetici i quali operando in analogia ai processi di evoluzione biologica, consentono di esplorare in maniera molto efficace lo spazio di ricerca [6-8]. La funzione di fitness oggetto dell'ottimizzazione è il profitto, al netto delle spese di acquisto e gestione, proveniente dall'operazione dell'impianto e valutato tramite la simulazione Monte Carlo sopra menzionata. Al fine di conseguire risultati statisticamente significativi in tempi di calcolo ragionevoli, per ogni potenziale soluzione proposta dall'algoritmo genetico vengono simulate solo poche centinaia di storie Monte Carlo e si sfrutta il fatto che le soluzioni 'buone' vengono selezionate più volte durante l'evoluzione 'naturale' della popolazione di soluzioni per "accumulare" risultati statisticamente significativi.

Nel prossimo capitolo viene illustrato il modello stocastico del sistema. Nel terzo capitolo vengono illustrati i concetti base degli algoritmi genetici. Il quarto capitolo propone l'impiego combinato di tecniche di simulazione Monte Carlo e algoritmi genetici per la soluzione di problemi di ottimizzazione complessi. L'applicazione del metodo proposto viene illustrata per mezzo di un caso studio.

2. ROTTURE, RIPARAZIONI, MANUTENZIONI, COSTI: IL MODELLO

Consideriamo un generico impianto schematizzabile in N unità operative (nodi) posti in serie (Figura 1).

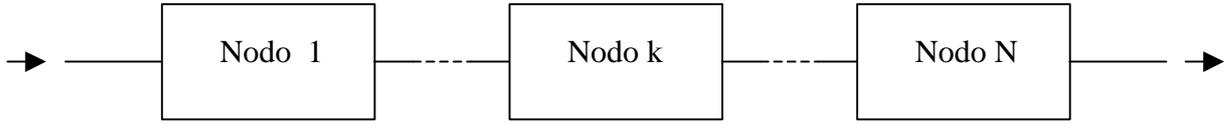


Figura 1 : schema dell'impianto

Il generico nodo k è formato da n_k componenti connessi tra loro secondo una determinata logica di funzionamento (serie, parallelo o standby). Si definisce stato del sistema al tempo t l'insieme degli stati di tutti i componenti che lo costituiscono. Si suppone che all'istante $t_0=0$ il sistema parta in condizioni nominali.

Si identificano inoltre gli stati di guasto del sistema (cut set). Tra questi si distinguono cut set di non servizio, per i quali il sistema non è in grado di fornire il servizio per cui è stato progettato, e cut set incidentali, che oltre a essere tali da impedire la fornitura del servizio causano conseguenze incidentali all'ambiente esterno. L'introduzione di quest'ultimo tipo di cut set permette di considerare 'impianti a rischio'.

Per quanto riguarda il comportamento di rottura dei componenti si ipotizza che le rotture avvengano a tempi stocastici governati da ratei variabili linearmente nel tempo [9]:

$$\lambda(t) = \lambda_0 + a \cdot t \quad (1)$$

dove I_0 è un termine costante e a è il fattore di invecchiamento.

La funzione di distribuzione cumulativa (cdf) dei tempi rottura del componente e la corrispondente densità di probabilità (pdf) sono pertanto:

$$F(t) = 1 - e^{-\int_0^t I(u) du} = 1 - e^{-\int_0^t (I_0 + a \cdot u) du} = 1 - e^{-(I_0 t + \frac{1}{2} a t^2)} \quad (2)$$

$$f(t) = \frac{dF(t)}{dt} = (I_0 + a \cdot t) \cdot e^{-(I_0 t + \frac{1}{2} a t^2)} \quad (3)$$

Ne consegue che il tempo medio di rottura (failure) è :

$$\bar{t}_f = \int_0^{\infty} t \cdot f(t) dt = \int_0^{\infty} (I_0 t + a t^2) e^{-(I_0 t + \frac{1}{2} a t^2)} dt = \frac{1}{2} \sqrt{\frac{2p}{a}} e^{\frac{I_0^2}{2a}} \left[1 - \Phi\left(\frac{I_0}{\sqrt{2a}}\right) \right] \quad (4)$$

dove $\Phi(x) = \frac{2}{\sqrt{p}} \int_0^x e^{-u^2} du$ è la funzione errore.

L'effetto dell'invecchiamento è mitigato da manutenzioni preventive di periodo t che contrastano la crescita di $\lambda(t)$ riportandola al valore λ_0 al termine di ciascun periodo di manutenzione. Affinchè le manutenzioni siano efficaci, t deve essere tale che $\lambda(t)$ non cresca molto in $[0, t]$. Per semplicità, al posto della cdf (2) utilizzeremo nelle nostre simulazioni la più semplice distribuzione esponenziale caratterizzata da un rateo di guasto I^* costante di valore tale da aversi la stessa probabilità di guasto in un periodo che si avrebbe se si usasse la distribuzione vera (2). Ne risulta che I^* deve avere il valore efficace (costante in tutto il periodo t):

$$\lambda^* = \frac{1}{t} \int_0^t I(t) dt = I_0 + a \cdot \frac{t}{2} \quad (5)$$

che rappresenta il valor medio della funzione $\lambda(t)$ sul periodo t . Ne nasce un evidente legame biunivoco tra I^* e t . Il tempo medio di rottura del componente può essere calcolato come :

$$\left(\bar{t}_f\right)_M = \frac{1}{I^*} = \frac{1}{I_0 + a \cdot \frac{t}{2}}$$

Anche per i tempi di riparazione si ipotizza una distribuzione di probabilità esponenziale. Sebbene il processo di riparazione sia tutt'altro che markoviano, questo assunto, che può essere facilmente rimosso in un approccio Montecarlo, permette di confrontare i risultati analitici con quelli ottenuti mediante simulazione Montecarlo. Inoltre, dal momento che la disponibilità asintotica di un componente riparabile dipende dal tempo medio di riparazione, l'approssimazione di rateo di riparazione costante dà effetti 'significativi' solo durante l'evoluzione transitoria [4].

Nella pratica diversi componenti possono richiedere diversi tipi di riparazione (elettronica, meccanica etc.). La singola riparazione è eseguita da squadre di operai specializzati. Ogni squadra può intervenire su un solo componente alla volta. Il numero delle squadre specializzate è generalmente inferiore a quello dei componenti che ne possono richiedere l'intervento. Al fine di una più razionale gestione degli interventi di riparazione, ad ogni componente viene assegnato un valore di priorità. I componenti con valori di priorità più alti saranno quelli su cui le squadre interverranno prima. Nel caso in cui tutte le squadre disponibili per un dato tipo di riparazione siano impegnate, la rottura di un componente con priorità superiore a quella di uno dei componenti in riparazione fa sì che venga interrotta la riparazione di quest'ultimo e la squadra vada ad occuparsi del nuovo componente rotto.

Si è introdotto inoltre il modello di Brown-Proschan [10] per tenere conto della possibilità che un componente non venga ben riparato cosicché le sue condizioni al termine della riparazione non siano quelle precedenti al guasto, ma bensì esso risulti più propenso alla rottura. Il modello prevede che a seguito di una riparazione imperfetta, che avviene con probabilità $q = 1 - p$:

- 1) i ratei di rottura del componente vengano amplificati tramite un fattore moltiplicativo p_1 (≥ 1) che rende le rotture più probabili;
- 2) i ratei di riparazione del componente vengano divisi per un fattore p_m (≥ 1) in quanto si ipotizza che la riparazione imperfetta rende il componente più difficile da riparare per cui la riparazione richiede un tempo mediamente più lungo.

Con probabilità p , invece, il componente viene riportato nelle condizioni precedenti il guasto (riparazione minima).

Per quanto riguarda la manutenzione periodica, il modello adottato si basa sulle seguenti ipotesi:

- i) la manutenzione viene eseguita con periodicità variabile a seguito degli effetti di riparazioni imperfette. In altre parole, viene prevista una strategia adattiva di manutenzione per far fronte al deterioramento del componente. Tale strategia stabilisce che a fronte di riparazioni imperfette che deteriorano il componente, si riducano i periodi di manutenzione in maniera da mantenere costante il rapporto fra il tempo medio di rottura $\frac{1}{I}$ e il periodo t (Figura 2). Dunque se la riparazione è minima (cioè perfetta), $I_{new} = I_{old}$ e $t_{new} = t_{old}$; al contrario, in seguito ad una riparazione imperfetta, $I_{new} = (1 + p_1) \cdot I_{old}$ e il periodo si modifica così che
$$t_{new} = t_{old} \cdot \frac{I_{old}}{I_{new}} = \frac{t_{old}}{1 + p_1};$$
- ii) al periodo t tra due manutenzioni successive contribuisce solo il tempo durante il quale il componente è attivo, mentre non contribuisce il tempo in cui esso è in stand by;
- iii) l'effetto delle manutenzioni è quello di riportare il componente nelle stesse condizioni in cui si trovava all'inizio del periodo precedente;
- iv) le azioni di manutenzione hanno durata nulla e sono sempre disponibili un numero sufficiente di tecnici per eseguirle;
- v) il periodo fra due manutenzioni è una frazione $\frac{1}{a_t}$ del tempo medio di guasto del componente:

$$\tau = \frac{1}{\alpha_\tau} \cdot t_f = \frac{1}{\alpha_\tau} \frac{1}{\lambda^*(\tau)} = \frac{1}{\alpha_\tau} \cdot \frac{1}{I_0 + a \cdot \frac{t}{2}} \quad (6)$$

$$\tau = \frac{\lambda_0}{a} \left[\sqrt{1 + \frac{1}{\alpha_\tau \frac{\lambda_0^2}{2a}}} - 1 \right] \quad (7)$$

Ne deriva un legame tra \mathbf{t} ed i parametri I_0 , a caratteristici del comportamento di rottura del componente.

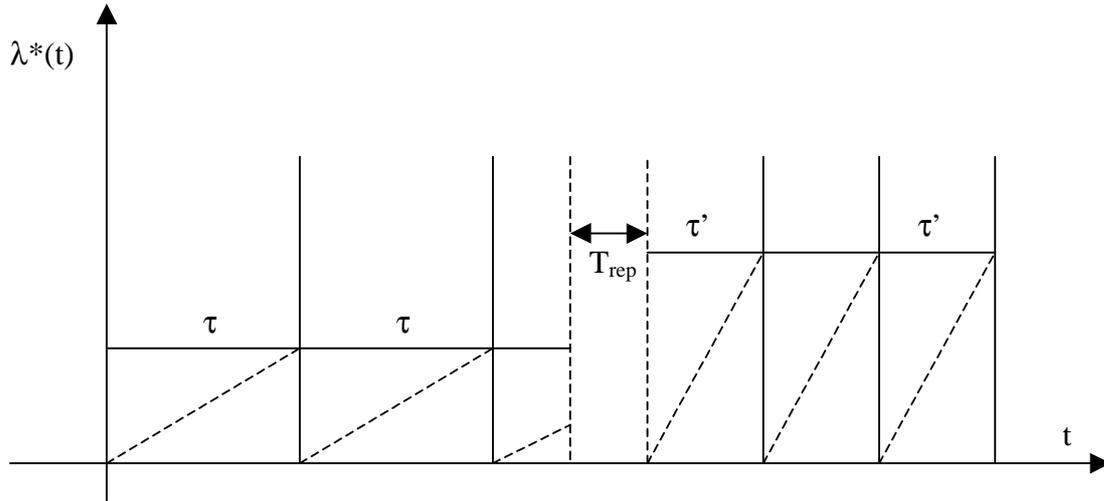


Figura 2: Dinamica dei periodi di manutenzione: a seguito di una riparazione imperfetta (di durata T_{rep}), il componente risulta deteriorato, così da richiedere una diminuzione del periodo di manutenzione da \mathbf{t} a \mathbf{t}' .

Per quanto riguarda gli aspetti economici faremo riferimento a costi e proventi di esercizio attualizzati all'istante di messa in funzione dell'impianto. Data una somma di denaro $S(t)$ disponibile al tempo discreto t misurato in anni, il corrispondente valore S_0 attualizzato al tempo $t_0 < t$ è dato dalla seguente espressione :

$$S_0 = \frac{S(t)}{(1+i)^{t-t_0}} \quad (8)$$

ove i è il tasso di attualizzazione che dipende da vari fattori economici. Per semplicità supporremo che esso sia costante durante tutta la vita dell'impianto.

L'obiettivo ultimo dell'attività di progettazione è quello di trovare la configurazione di sistema che consenta di ottenere il massimo profitto dall'esercizio del sistema, al netto delle spese. Dal punto di vista matematico ciò si traduce in un problema di ricerca del massimo della funzione profitto netto, G , così definita:

$$G = I - (C_{AI} + C_R + C_{NS} + C_{INC}) \quad (9)$$

Il primo termine a secondo membro rappresenta il ricavato utile del sistema,

$$I = I_t \cdot \int_0^{T_M} \frac{A(t)dt}{(1+i)^t} \quad (10)$$

dove :

I_t = ricavo utile per unità di tempo (incasso – spese fisse) supposto costante. L'incasso rappresenta la somma di denaro per unità di tempo che viene pagata al gestore dell'impianto dal cliente a cui viene venduto il servizio. Le spese fisse sono un termine di flusso di cassa uscente per unità di tempo presente nel bilancio sia nel caso in cui l'impianto funzioni sia che non funzioni: esse sono associate allo stipendio delle squadre di riparazione, alle operazioni di manutenzione etc.

$A(t)$ = disponibilità dell'impianto all'istante t

T_M = tempo di missione del sistema

Il secondo termine a secondo membro della (9), C_{AI} , rappresenta il costo di acquisto e installazione di tutti i componenti,

$$C_{AI} = \sum_{j=1}^{N_C} C_j \quad (11)$$

dove :

C_j = costo di acquisto e installazione del componente j

N_C = numero totale di componenti del sistema

Si immagina che l'acquisto e installazione dell'impianto vengano fatti con denaro preso a prestito all'istante iniziale $t_0=0$ e da restituirsi con tasso di interesse i . La restituzione di tale prestito avviene

in rate annuali costanti di valore nominale $C_{AI} \cdot \left(i + \frac{i}{(1+i)^{T_M} - 1} \right)$. Nel caso in cui l'impianto

debba essere chiuso prima dell'esaurirsi della sua vita (per esempio a causa di un incidente che provoca un danno irreparabile) le rate residue vengono pagate subito al fine di evitare di pagare gli interessi nel seguito. Si suppone inoltre che il valore residuo dell'impianto a fine vita (T_M o chiusura prematura) sia nullo.

Il terzo termine a secondo membro della (9), C_R , rappresenta i costi di riparazione dei componenti del sistema,

$$C_R = \sum_{j=1}^{N_C} C_{Rj} \cdot \int_0^{T_M} \frac{I_{Rj}(t) dt}{(1+i)^t} \quad (12)$$

dove :

C_{Rj} = costo di riparazione per unità di tempo del componente j , supposto costante nel tempo

$I_{Rj}(t)$ = variabile indicatore della riparazione del componente j al tempo t

$$= \begin{cases} 1 & \text{se il componente è in riparazione all'istante } t \\ 0 & \text{viceversa} \end{cases}$$

Il quarto termine, C_{NS} , rappresenta il costo totale per mancata fornitura del servizio,

$$C_{NS} = C_U \cdot \int_0^{T_M} \frac{[1 - A(t)] \cdot I_{NS}(t) dt}{(1+i)^t} \quad (13)$$

dove :

C_U = costo di non servizio per unità di tempo (inclusivo di penale e spese fisse di gestione e diminuito dei costi variabili legati al servizio), supposto costante. La penale rappresenta la somma di denaro che deve essere pagata al cliente come risarcimento per il mancato servizio. Nel caso in cui l'erogazione del servizio termini prematuramente a causa di un incidente grave e irreparabile si suppone che la penale venga pagata al massimo per una frazione fissata di T_M ; trascorso tale tempo il cliente viene considerato completamente risarcito per il danno subito a causa del mancato servizio

$I_{NS}(t)$ = variabile indicatore del pagamento della penale al tempo t

$$= \begin{cases} 0 & \text{se si è verificato un incidente grave il cui risarcimento è terminato} \\ 1 & \text{altrimenti} \end{cases}$$

L'ultimo termine nella (9) tiene conto dei costi associati alle conseguenze di incidenti gravi,

$$C_{INC} = \sum_{k=1}^{N_{INC}} I_{INC,k} \cdot \frac{C_{INC,k}}{(1+i)^{t_{INC,k}}} \quad (14)$$

dove :

N_{INC} = numero di tipi di incidente; si ipotizza che ogni incidente sia irreparabile per cui il corrispondente cut set di sistema rappresenta uno stato assorbente

$C_{INC,k}$ = costo associato al k-esimo incidente

$I_{INC,k}$ = variabile indicatore del verificarsi dell'incidente di tipo k

$$= \begin{cases} 1 & \text{se l'incidente di tipo k si è verificato} \\ 0 & \text{altrimenti} \end{cases}$$

$t_{INC,k}$ = anno in cui si verifica l'incidente di tipo k

3. ALCUNE NOZIONI BASE SUGLI ALGORITMI GENETICI

Gli algoritmi genetici (che nel seguito indicheremo con la sigla GA, dall'inglese 'Genetic Algorithms') sono strumenti numerici per la ricerca del massimo di una assegnata funzione a valori reali e non negativa. Tale funzione, detta funzione obiettivo o fitness, può dipendere da una o più variabili, eventualmente soggette a vincoli lineari o non lineari. Il nome stesso di tali algoritmi descrive il loro modo di funzionamento: ispirandosi alle leggi della selezione naturale, essi fanno emergere la soluzione del problema selezionandola tra un gran numero di soluzioni 'candidate', che vengono accoppiate tra loro, esattamente come avviene in natura, in un processo 'evolutivo' che favorisce la sopravvivenza delle soluzioni migliori.

A seguito dell'analogia con i processi di selezione naturale, la terminologia usata nei GA rispecchia, a grandi linee, quella usata in biologia. Per questo, nel problema di massimizzazione di una funzione, le soluzioni "candidate" vengono chiamate 'cromosomi' e sono formate da una stringa di variabili, proprio come in biologia gli individui sono identificati da cromosomi, formati da sequenze di basi (Adenina-Guanina-Citosina-Timina); il cromosoma è suddiviso in 'geni' che formano il genotipo e la cui decodifica in numeri reali, chiamati fattori di controllo (uno per ciascun gene), costituisce il fenotipo del cromosoma.

I fattori di controllo rappresentano i valori delle variabili da cui dipende la funzione obiettivo e consentono perciò la valutazione di tale funzione in corrispondenza del determinato cromosoma. Il valore della funzione obiettivo così ottenuto è detto 'fitness' del cromosoma in esame e dà un'indicazione della bontà della soluzione codificata nel cromosoma. In altre parole, considerando gli n fattori di controllo di un individuo come variabili indipendenti in un iperspazio n-dimensionale, una popolazione può essere vista come una collezione di punti in esso; a ciascuno di tali punti è associato un numero, la fitness dell'individuo corrispondente, che ne quantifica la 'bontà' come soluzione del problema di ottimizzazione in esame.

La ricerca della soluzione da parte del GA avviene attraverso l'esame di successive generazioni di popolazioni di cromosomi che evolvono in maniera tale che, possibilmente, ogni generazione sia migliore della precedente in termini di fitness. Il metodo che consente di creare una generazione a partire dalla precedente rispecchia il processo naturale della riproduzione: tra gli individui della popolazione corrente viene scelta una coppia di genitori che vengono fatti accoppiare (i loro cromosomi vengono incrociati) dando luogo così a due figli. Dei quattro individui ora a disposizione (due genitori e due figli) due soli sopravvivono a seconda del loro livello di fitness. Il processo di accoppiamento, ripetuto per più coppie di genitori, e l'altrettanto naturale fenomeno della mutazione genetica daranno origine a una nuova popolazione che, a sua volta, si riprodurrà: dopo un certo numero di generazioni si arriverà a una popolazione che, sperabilmente, conterrà, in almeno un suo individuo, la soluzione ottimale al problema affrontato.

Il campo di potenziale applicabilità dei GA è molto vasto [7]: ciò è dovuto al fatto che, grazie ad alcune loro proprietà intrinseche, essi sono in grado di risolvere in maniera efficace problemi difficilmente affrontabili con metodi classici di ottimizzazione. Un primo importante vantaggio rispetto ai metodi classici consiste nel fatto che i GA non risentono del numero di variabili come i metodi numerici classici, il cui peso computazionale e la cui imprecisione (dovuta essenzialmente ad errori di arrotondamento) aumentano rapidamente all'aumentare del numero di variabili. Pertanto i GA sono spesso utilizzati in problemi in cui la funzione obiettivo dipende da un grande numero di variabili (tipicamente 5 o più). Una seconda proprietà di questi algoritmi è che il metodo di ricerca della soluzione non presuppone alcun tipo di linearità del problema, per cui è possibile affrontare anche problemi in cui esistono una o più non linearità, eventualmente anche sotto forma di vincoli. Un ulteriore vantaggio rispetto ai metodi classici è che il metodo di ricerca dei GA si basa sulla conoscenza della sola funzione obiettivo, al contrario dei metodi classici che necessitano anche di altre informazioni, quali le derivate della funzione etc. Il GA risulta quindi particolarmente utile

quando si ha a che fare con funzioni discontinue, o date per punti, e di cui non si conosce nessuna informazione sull'andamento.

4. L'UTILIZZO COMBINATO DI MONTE CARLO ED ALGORITMI GENETICI IN PROBLEMI DI OTTIMIZZAZIONE

La procedura degli algoritmi genetici prevede che durante l'evoluzione generazionale delle popolazioni, in corrispondenza di ogni soluzione proposta si valuti la funzione di fitness. Dato il vasto spazio di ricerca, il numero di soluzioni considerato è tipicamente molto elevato e ciò rende impensabile l'esecuzione di una simulazione Monte Carlo completa (migliaia di storie) per ogni cromosoma analizzato.

Per ovviare a questo problema, durante l'evoluzione della popolazione viene mantenuto aggiornato un archivio delle configurazioni migliori tra quelle via via proposte dall'algoritmo, e dei corrispondenti valori di funzione obiettivo, profitto, costi e disponibilità. Quando l'archivio è pieno la nuova configurazione va a sostituire quella dell'archivio con fitness minima. Quando durante l'evoluzione una configurazione viene riproposta più volte, i risultati ad essa relativi vengono dinamicamente aggiornati.

Questo modo di procedere consente di ottenere stime Montecarlo con statistica tanto migliore quanto più alta è la fitness della configurazione, cioè quanto migliore è la soluzione proposta. Infatti il numero di volte in cui viene chiamata una certa configurazione durante l'evoluzione della popolazione dipende in parte dalla sua fitness: cromosomi con fitness maggiore sopravvivono e vengono selezionati come genitori con maggiore probabilità di cromosomi con fitness minore. Ne consegue che il numero di storie eseguite tende a essere in media maggiore per le configurazioni migliori (fitness maggiore): la stima delle grandezze di interesse per queste configurazioni risulta pertanto avere una statistica migliore. E' chiaro allora che per ogni chiamata di una configurazione è sufficiente eseguire relativamente poche storie Montecarlo (ad es. 100 nelle nostre applicazioni) cosicché le configurazioni lontane dall'ottimo vengono esaminate poche volte, evitando così di sprecare tempo calcolo su configurazioni 'perdenti', mentre per la configurazione ottimale, e quelle più vicine ad essa, vengono eseguite molte più storie che consentono di ottenere stime statisticamente significative.

5. OTTIMIZZAZIONE DEL PROGETTO DI UN IMPIANTO: UN CASO STUDIO

Il problema in esame consiste nella scelta ottimale della componentistica e del layout di un impianto complesso. Tale scelta è dettata dall'interazione tra esigenze economiche, di costo e guadagno, ed esigenze affidabilistiche e di sicurezza nei confronti di incidenti con conseguenze gravi sull'ambiente. Lo scopo della progettazione è dunque quello di individuare la configurazione di sistema che consente di ottenere il maggior provento economico netto medio in ottemperanza con i necessari requisiti di sicurezza.

Come precedentemente detto il sistema viene definito tramite un certo numero N di unità operative (nodi) posti in serie. La definizione delle possibili configurazioni alternative del sistema necessita :

- la definizione dei possibili tipi di componenti utilizzabili nei vari nodi, con le loro caratteristiche di rottura e riparazione,
- la definizione delle possibili configurazioni c_n di ciascun nodo n, $n=1,2,\dots,N$.

La configurazione funzionale del sistema viene poi completata da informazioni relative a :

- tipi di riparazione e numero di squadre disponibili per ciascuna di esse,
- logiche di funzionamento per ogni nodo,
- configurazioni di cut set per ogni nodo.

Al fine di utilizzare il GA per trovare la configurazione ottimale tra quelle proposte è necessario stabilire una regola di corrispondenza biunivoca tra le configurazioni del sistema e i cromosomi. Ogni configurazione di ciascun nodo viene etichettata da un indice intero: dal momento che la ricerca della configurazione ottimale viene eseguita scegliendo separatamente la configurazione di ogni singolo nodo, il cromosoma deve contenere un numero N di indici pari al numero N di nodi di cui è composto il sistema. Per la codifica, si è scelto di prendere un cromosoma formato da un unico gene contenente gli indici delle configurazioni di tutti i

nodi in una stringa di $k = \sum_{n=1}^N k_n$ bit, ove k_n è il numero minimo di bit necessario per rappresentare le c_n

configurazioni del nodo n. Ad esempio, un nodo che può assumere 6 configurazioni diverse verrà rappresentato da una stringa costituita da 3 bit (Tabella 1).

Indice della configurazione	Stringa corrispondente
1	000
2	001
3	010
4	011
5	100
6	101

Tabella 1 : codifica di 6 configurazioni di un nodo

I cromosomi che codificano indici di configurazioni inesistenti vengono scartati prima del loro ingresso nella popolazione. Con riferimento all'esempio precedente, se un accoppiamento o una mutazione creassero un individuo con configurazione di indice 7 o 8 esso verrebbe eliminato automaticamente.

La scelta di questo tipo di codifica fa sì, rispetto ad una codifica con un gene per nodo, che dal crossover vengano generati figli con tutti i nodi uguali a quelli dei genitori eccetto quello in corrispondenza del quale si va a collocare il separatore. Questo modo di procedere evita una eccessiva dispersione del patrimonio genetico, che potrebbe rallentare o impedire la convergenza, e contemporaneamente favorisce la sopravvivenza delle configurazioni migliori, consentendo una più rapida convergenza.

Per quanto riguarda l'evoluzione della popolazione, in questo lavoro la fase di selezione dei genitori per l'accoppiamento è eseguita mediante la regola della roulette, cioè scegliendo i genitori con probabilità proporzionali ai rispettivi valori di fitness; l'accoppiamento o crossover viene ottenuto mediante inserimento casuale di un separatore nei geni omologhi dei genitori selezionati; la sostituzione nella popolazione prevede che dei quattro cromosomi che derivano dall'accoppiamento (due genitori e due figli) sopravvivano i due 'migliori', cioè con fitness più grande; infine, la mutazione genetica avviene con probabilità 10^{-3} .

La funzione guadagno, cioè la fitness, corrispondente ad una configurazione proposta, viene calcolata per mezzo del codice Montecarlo UNRELMC 1.0 [11] sviluppato precedentemente nel nostro Dipartimento e opportunamente modificato. Per quanto detto alla Sezione 4, alla prima proposta di una data configurazione vengono eseguite 1000 storie mentre per ogni chiamata successiva se ne eseguono solamente 100, che vanno via via irrobustendo la statistica dei risultati ottenuti per quella configurazione.

Per assegnate caratteristiche di componentistica, logica di funzionamento/guasto (descritta per mezzo dell'introduzione di opportune connessioni secondo lo schema utilizzato in (12) e tempo di missione, il codice Montecarlo (che nel seguito indicheremo con la sigla MC) simula un prefissato numero di storie di evoluzione del sistema e per ognuna calcola le grandezze:

- $A(t)$ = disponibilità del sistema
- T_S = tempo di servizio del sistema 'attualizzato'
- T_{NS} = tempo di non servizio del sistema 'attualizzato'
- tempo di riparazione 'attualizzato' per ogni componente

e registra il verificarsi di un eventuale incidente grave. L'aver supposto costante il ricavo e i costi per unità di tempo permette di 'attualizzare' i tempi invece che i flussi di cassa, consentendo un notevole risparmio dal punto di vista computazionale.

Al termine di un prefissato numero di storie vengono calcolati i valori medi che costituiscono le stime Montecarlo delle grandezze di cui sopra. Dalla stima dei proventi di esercizio e dai costi di acquisto, riparazione, non servizio e incidente è possibile calcolare il guadagno medio per la fissata configurazione dell'impianto secondo la (9).

Nella nostra applicazione l'impianto considerato costituisce un esempio tipico di impianto di processamento di materiale greggio [13]. Il sistema può essere schematizzato da cinque nodi in serie, ciascuno dei quali preposto all'adempimento di un particolare passo del processo di frantumazione.

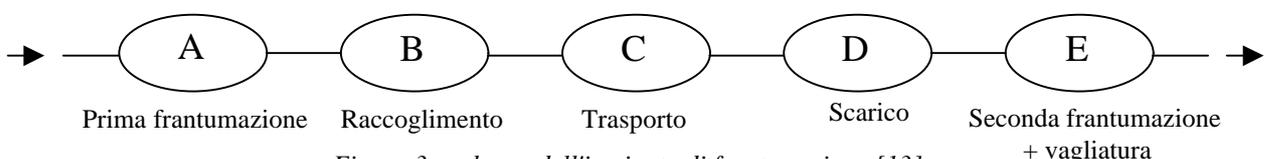


Figura 3 : schema dell'impianto di frantumazione [13].

Per ogni nodo si possono avere diverse scelte progettuali come indicato in Tabella 2. Il numero totale di possibili configurazioni di sistema risulta pertanto essere $7 \cdot 16 \cdot 14 \cdot 14 \cdot 7 = 153664$.

Le configurazioni del sistema sono pertanto rappresentabili per mezzo di un unico gene strutturato nel seguente modo :

- nodo A: 7 configurazioni → 3 bit
- nodo B: 16 configurazioni → 4 bit
- nodo C: 14 configurazioni → 4 bit
- nodo D: 14 configurazioni → 4 bit
- nodo E: 7 configurazioni → 3 bit

per un totale di 18 bit. Il range del gene è tale che il fattore di controllo associato a ciascuna configurazione sia un numero intero.

Il grande spazio di ricerca ha reso necessario l'utilizzo di una popolazione molto vasta (500 individui); si è osservato, in compenso, che la soluzione ottimale viene raggiunta in un numero ridotto di generazioni (25).

Nodi	Numero di configurazioni alternative	Tipi di componenti	Logica
A	7	a	3-out-of-3 G 3-out-of-4 G 3-out-of-5 G 3-out-of-6 G 3-out-of-7 G 3-out-of-8 G 3-out-of-9 G
B	16	b1, b2, b3	2-out-of-2 G 2-out-of-3 G
C	14	c1, c2	1-out-of-1 G 1-out-of-1 G + 1 standby 1-out-of-1 G + 2 standby
D	14	d1, d2	1-out-of-1 G 1-out-of-1 G + 1 standby 1-out-of-1 G + 2 standby
E	7	e	3-out-of-3 G 3-out-of-4 G 3-out-of-5 G 3-out-of-6 G 3-out-of-7 G 3-out-of-8 G 3-out-of-9 G

Tabella 2: configurazione dei potenziali di nodo

I valori numerici dei ratei di rottura e riparazione sono stati presi dall'impianto reale [13], opportunamente modificati (Tabella 3). Inoltre, si sono fatte le seguenti ipotesi per quanto riguarda l'invecchiamento dei componenti:

- probabilità di riparazione imperfetta = 0.3
- fattore moltiplicativo dei ratei di rottura $p_1 = 1.5$
- fattore riduttivo dei ratei di riparazione $p_m = 1.3$

Al fine di considerare il sistema un impianto a rischio si definiscono due tipi di incidenti gravi: essi corrispondono alla rottura dei nodi di prima e seconda frantumazione A ed E.

Si suppone inoltre che esistano due tipi differenti di riparazioni, il primo relativo alle unità di frantumazione (nodi A e E) e il secondo relativo alle unità di movimentazione del materiale (nodi B, C e D); il numero di squadre di ciascun tipo è sufficiente alla riparazione contemporanea di tutti i componenti. Si fa infine l'ipotesi che i componenti non si possano rompere mentre sono in standby. I dati relativi ai componenti e al sistema sono riportati nelle Tabella 3 e 4.

Componente i	Rateo di rottura λ_i [y ⁻¹]	Rateo di riparazione μ_i [y ⁻¹]	Costo d'acquisto C_i [10 ⁶ \$]	Costo di riparazione C_{Ri} [10 ⁶ \$.y ⁻¹]
a	$1.5 \cdot 10^{-2}$	$4.0 \cdot 10^{-2}$	3.0	0.55
b1	$2.0 \cdot 10^{-4}$	$8.0 \cdot 10^{-3}$	5.0	10.0
b2	$2.0 \cdot 10^{-3}$	$8.0 \cdot 10^{-2}$	3.0	6.2
b3	$2.0 \cdot 10^{-2}$	$8.0 \cdot 10^{-1}$	1.0	2.1
c1	$1.0 \cdot 10^{-4}$	$8.0 \cdot 10^{-3}$	10.0	41.0
c2	$1.0 \cdot 10^{-3}$	$8.0 \cdot 10^{-2}$	5.0	20.0
d1	$1.0 \cdot 10^{-4}$	$8.0 \cdot 10^{-3}$	7.0	28.0
d2	$1.0 \cdot 10^{-3}$	$8.0 \cdot 10^{-2}$	3.0	12.0
E	$1.7 \cdot 10^{-2}$	$4.0 \cdot 10^{-2}$	5.0	0.85

Tabella 3 : parametri dei componenti

Incasso I_t	[10 ⁶ \$.y ⁻¹]	20.0
Costo di non servizio C_U	[10 ⁶ \$.y ⁻¹]	20.0
Costo d'incidente $C_{INC,1}$	[10 ⁶ \$]	50.0
Costo d'incidente $C_{INC,2}$	[10 ⁶ \$]	70.0
Tasso d'interesse annuo i		3%
Tempo di missione T_M	[y]	50
Tempo massimo in cui pagare la penale	[y]	5

Tabella 4 : parametri del sistema

I valori scelti consentono di fare osservazioni qualitative a priori sulle caratteristiche della configurazione ottimale del sistema: in tale modo sarà possibile una validazione parziale della procedura di ricerca.

I nodi A ed E sono caratterizzati da componenti uguali e in parallelo tra loro: dal momento che tali nodi (a cui sono stati associati costi appositamente elevati) sono gli unici responsabili degli incidenti, per essi la scelta ottimale ricadrà su configurazioni di elevata affidabilità, al limite quella con il maggior numero di componenti ridondanti (3/5).

I componenti utilizzabili nei nodi B, C e D sono caratterizzati da costi d'acquisto molto differenti tra loro: al fine di minimizzare il costo totale d'acquisto dell'impianto è pertanto opportuno scegliere componenti poco costosi. Tale scelta deve però essere compatibile con l'esigenza di minimizzare il tempo (e il corrispondente costo) di non servizio: nel caso del nodo B tale risultato si potrebbe ottenere con l'introduzione di una ridondanza (configurazione 2/3); nel caso dei nodi C e D i componenti con costo minore hanno vita media pari a $20 \cdot T_M$ e di conseguenza garantiscono un'affidabilità sufficiente senza bisogno di aggiungere alcun componente in standby.

Sulla base di queste osservazioni la configurazione ottimale a priori prevedibile potrebbe essere la seguente (Figura 4):

- nodo A : 5 frantumatori 'a' in parallelo con logica 3/5
- nodo B : 3 raccoglitori in parallelo con logica 2/3
- nodo C : 1 trasportatore 'c2'
- nodo D : 1 scaricatore 'd2'
- nodo E : 5 frantumatori+vagliatori 'e' in parallelo con logica 3/5

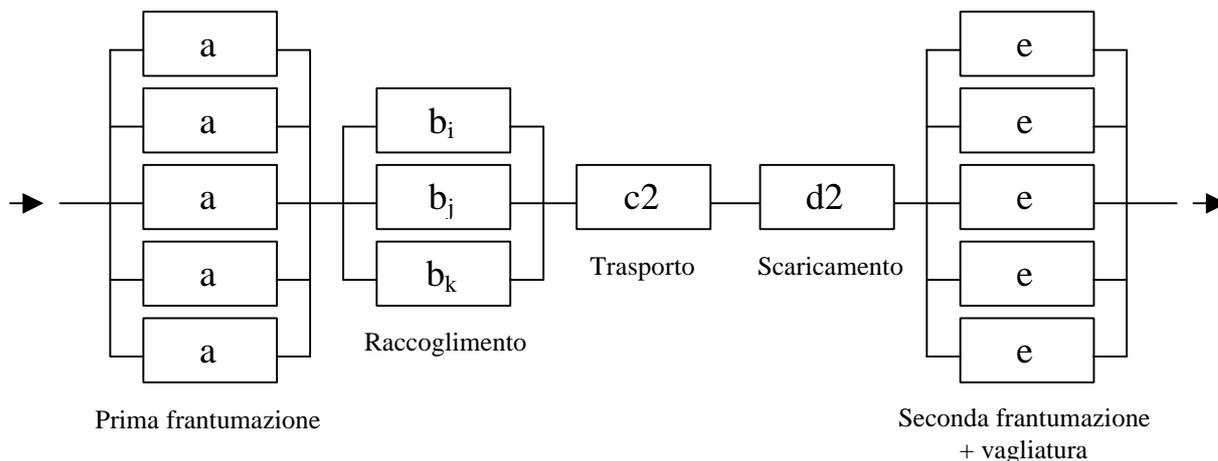


Figura 4 : configurazione ottimale dell'impianto

La ricerca della soluzione per mezzo del GA è stata da noi eseguita in due passi successivi.

Il primo passo è consistito nel cercare la configurazione ottimale tra le 153664 possibili. L'esame dell'archivio delle configurazioni a convergenza evidenzia che le configurazioni del sistema aventi fitness (guadagno) maggiore hanno tutte i nodi A ed E costituiti da 5 elementi con logica 3/5, corrispondenti alla configurazione ottimale da noi prevista in precedenza. La Figura 5 riporta l'andamento della fitness delle configurazioni migliori a convergenza: fino alla 676-esima i nodi A ed E hanno logica 3/5, dopodiché il guadagno ha una decrescita dapprima rapida e poi più dolce, dovuta all'introduzione in tali nodi di configurazioni diverse dalla 3/5.

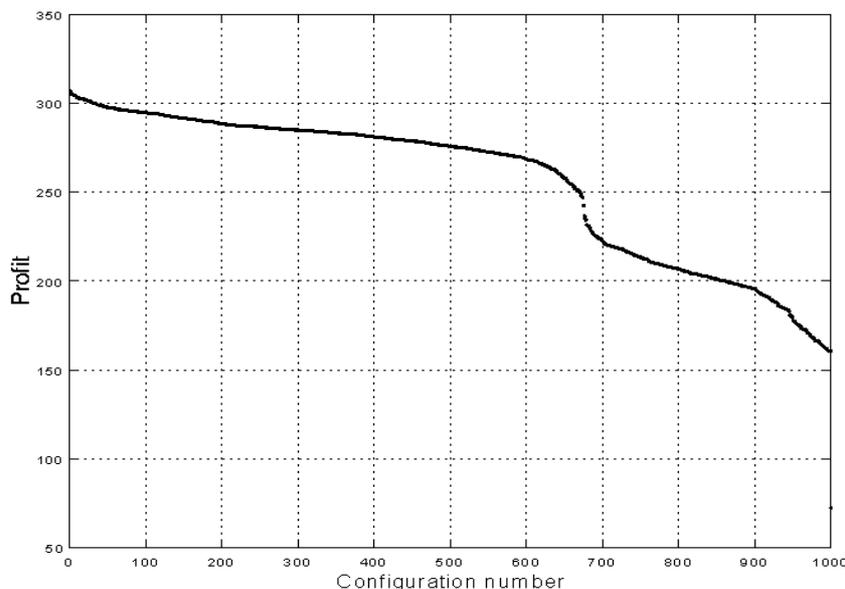


Figura 5 : fitness delle 1000 configurazioni migliori

In base a tale osservazione abbiamo deciso di raffinare la ricerca imponendo a priori le configurazioni 3/5 per i nodi A ed E: questo consente di ridurre lo spazio di ricerca a sole $16 \cdot 14 \cdot 14 = 3136$ configurazioni possibili. In questo caso è sufficiente un cromosoma di 12 bit (4 bit per nodo).

La nuova ricerca tramite l'algoritmo genetico porta alla configurazione ottimale di Figura 4, cioè proprio a quella da noi dedotta in precedenza, con $b_i = b_j = b_1$ e $b_k = b_3$. A conferma della robustezza dell'approccio, si è verificato che diversi valori del seme iniziale della catena random portano a questa stessa configurazione ottimale.

Dall'ordinamento dell'archivio si osserva che la soluzione ottimale risulta migliore della seconda di una quantità dieci volte più grande della corrispondente deviazione standard. Come ulteriore controllo sono stati eseguiti tre calcoli Monte Carlo con 10^6 storie per le tre migliori configurazioni: i risultati riportati in Tabella

5 confermano che la soluzione ottimale è migliore delle altre al di là della dispersione statistica legata alla stima Monte Carlo.

Configurazioni dell'archivio	Guadagno calcolato dal MC con 10⁶ storie
1 (ottimale)	471.57 ± 0.08
2 (prima subottimale)	470.20 ± 0.05
3 (seconda subottimale)	469.39 ± 0.07

Tabella 5 : configurazioni migliori

6. CONCLUSIONI:

Nella progettazione e gestione di un impianto a rischio all'analisi di sicurezza occorre affiancare quella dei costi, in considerazione di importanti fattori quali il mancato servizio, gli interventi di riparazione, il possibile verificarsi di incidenti gravi, l'invecchiamento dei componenti.

Questi aspetti complicano notevolmente il problema della ricerca della configurazione e gestione di sistema ottimale, cosicché una trattazione analitica non è facilmente percorribile.

In questo lavoro viene proposta la combinazione di algoritmi genetici di ricerca e tecniche di simulazione Monte Carlo per la soluzione di problemi di ottimizzazione riguardanti il progetto e la gestione di impianti complessi. L'algoritmo genetico considera l'evoluzione di una popolazione di cromosomi che codificano le diverse soluzioni alternative. In corrispondenza di ogni soluzione proposta, la simulazione Monte Carlo ne valuta la 'bontà' in termini di una predefinita funzione obiettivo, il guadagno, che rappresenta la funzione da massimizzare nella ricerca. Al fine di limitare il numero di storie Monte Carlo da simulare, è stato introdotto un metodo di raccolta dei risultati che consente di ottenere stime statisticamente significative per quelle soluzioni che, in quanto migliori, si ripresentano spesso durante l'evoluzione delle generazioni.

L'applicazione della procedura alla ricerca della configurazione ottimale di un impianto di processo di materiale greggio ne ha dimostrato l'efficienza a fronte dell'enorme spazio di ricerca (più di 150000 possibili soluzioni).

REFERENZE

- [1] Fletcher, R., Practical Methods of Optimization. John Wiley and Sons, NY, 1987.
- [2] Hilsmeier, T.A., Aldemir, T. and Vesely, W.E., Time-dependent unavailability of aging standby components based on nuclear plant data. Reliab. Engng. and Sys. Safety, **47**, (1995), pp. 199-205.
- [3] Martorell, S., Munoz, A. and Serradell, V., An approach to integrating surveillance and maintenance tasks to prevent the dominant failure causes of critical components, Reliab. Engng. and Sys. Safety, **50**, (1995), pp. 179-187.
- [4] Vaurio, J.K., On time-dependent availability and maintenance optimization of standby units under various maintenance policies, Reliab. Engng. and Sys. Safety, **56**, (1997), pp. 79-89.
- [5] Marseguerra, M. and Zio, E., Monte Carlo Evaluation of Corrective and Preventive Maintenance Strategies, Proceedings of ESREL2000, May 14-17, 2000, Edimburgh, Scotland, UK, pp. 269-274.
- [6] Holland, J.H., Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, MI, 1975
- [7] Goldberg, D.E., Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Publishing Company, 1989.
- [8] Chambers, L., Pratical Handbook of Genetic Algorithms, Vol. 1 and 2, 1995, CRC Press.
- [9] NUREG/CR-4769, Risk Evaluations of Aging Phenomena: The Linear Aging Reliability Model and Its Extension, US Nuclear Regulatory Commission, April 1987.
- [10] Brown, M. and Proshan, F., Imperfect Repair, Journal of Applied Probability, **20**, (1983), pp. 851-859..
- [11] Marseguerra M. and Zio, "UNRELMC 1.0: Un codice Monte Carlo per l'Analisi dell'Affidabilità Statica o Dinamica di un Sistema", Atti del Convegno Nazionale per la Valutazione e Gestione del Rischio negli Insediamenti Civili ed Industriali, 6-8 Ottobre 1998, Pisa, Italy, CD-ROM
- [12] Marseguerra, M. and Zio, E., Nonlinear Monte Carlo reliability analysis with biasing towards top event, Reliab. Engng. and Sys. Safety, **40**, (1993), pp. 31-42.
- [13] Henley E.J. e Kumamoto H., Probabilistic Risk Assessment, pagg. 554-561, IEEE Press